Che-Square Test

-) Chi-Square test is a statistical test used to determine whether there is a significant association by we categorical variables. It is positivularly useful for analyzing data where you have two or more ategorical variables and you want to assess whether they we independent or related in some way. There are several variation of this square test.

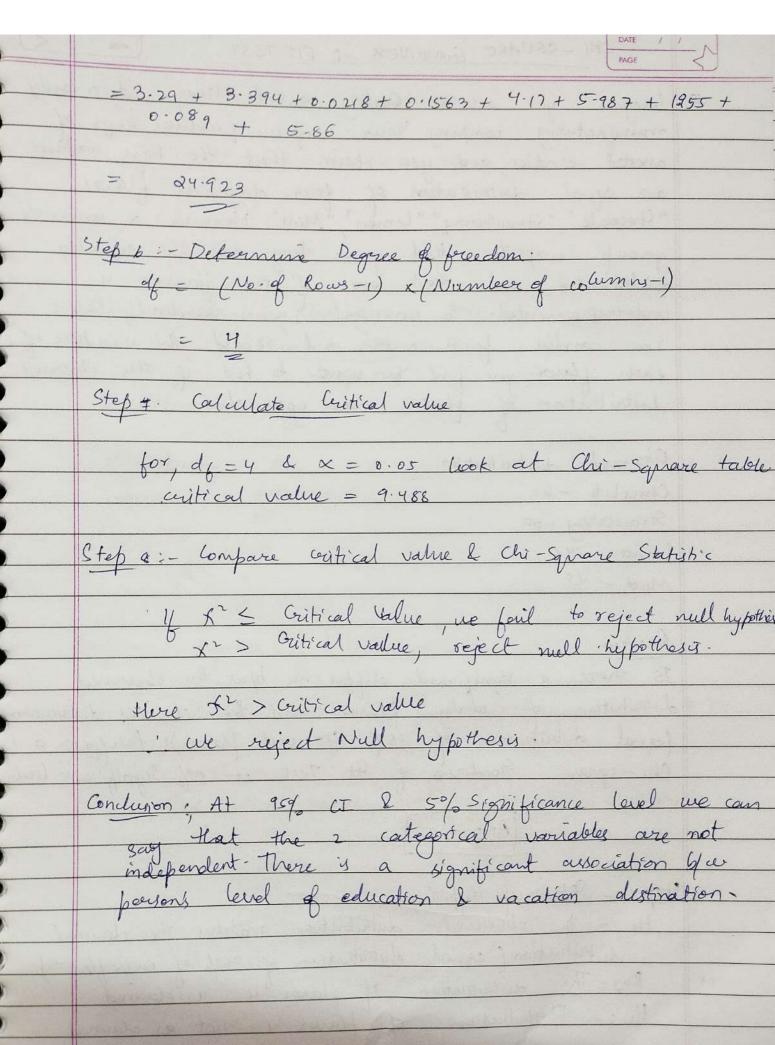
chi-Square tolde for Independence: - This test is used to determine if there is significant association by two categorical variables in contingency table. It helps answer question like "is there a relationship by gender and voting preference."

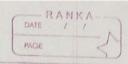
cohether an observed frequency distribution of categorical data matches an expected theoretical distribution. Go eg, you might we it to check if the distribution of blood types follows expected proportions.

to compare the distributions of categorical variables across different groups or populations to see if they are significantly different.

-	CHI- SOUARE TEST FOR MISE
	Suppose you are conducting a survey to investigate whether there is an explation you a person's level of education & their pereferred type of vacation destination. You collect data from 300 individuals and categorie them into 3 education levels ("High School", "Presholos Degree", and "Masters Degree") and three vacation pereferences ("Beach" "City" "Mountain"). Your goal is to determine there is a significant association by education level and vacation preference.
3	Vacation Preference
3	Education level Beach City Mountain
3	High School 50 30 20
3	Badrelos's Jegree 40 60 10 Masters Degree 30 40 30
i)	Is there a significant association blu a person's loud of education & their preferred type of vacation destination? Conduct a Chi-square test for Independence at a 5% significance level.
2	ANSWER
	Step 2: Set up typothesis
9 -	Null Hypothesis (Ho): There is no association b/w education level & vacation pereference.
	Ho:- Two categorical variables are independent. Attennative Hypothesis (H.): There is an association blu education level and vacation preference.
	14:- The two categorical variables are not independent.

5	PAGE / /	
1 15	Step 2: Choosing Significan level.	
la usi	Step 2: Choosing Significan level.	
His	the state of the s	
n.sk	Step 3: - Create contigence table. (Look at the table)	
. The	(book at the table)	#
2743	The state of the s	
00.7	Step 4: - Calculate Expected frequencies	
	Expected teramency = Row for to 1 x Column 1040	
	Expected frequency = Row total x Column total Grand total	
	the designation of the section of th	
	Education Level Beach City Mountain Total	
	High School 50 39 20 => 100	3
	Bachelors Degree 40 60 10 => 110	
	Masters Degree 30 40 30 => 100	
	120 130 cm 60 310	
	V=N 1 1 C . Grand total	
THAT	Expected frequencies Grand total	
ing.	A Service of the serv	
ilser artis	Education level Beach City Mountain	
iser arta	Education level Beach City Mountain	
isw arts	Education level Beach City Mountain	
ilver actual	Education level Beach City Mountain. High School 38.71 41.93 19.35	
iner arts	Education level Brach City Mountain High School 38.71 41.93 19.35 Bachelor's Degree 42:58 46.129 21.29 Mayter's Degree 38.7 41.935 19.35	
Degrada de la companya della companya della companya de la companya de la companya della company	Education level Brack City Mountain High School 38.71 41.93 19.35 Backelor's Degree 42:58 46.129 21.29 Mayter's Degree 38.7 41.935 19.35 Stop 5:- Calculate Chi-square Statistic	
	Education level Brack City Mountain High School 38.71 41.93 19.35 Backelor's Degree 42:58 46.129 21.29 Mayter's Degree 38.7 41.935 19.35 Stop 5:- Calculate Chi-square Statistic	
	Education level Brack City Mountain High School 38.71 41.93 19.35 Backelor's Degree 42:58 46.129 21.29 Mayter's Degree 38.7 41.935 19.35 Stop 5:- Calculate Chi-square Statistic	
	Education level Beach City Mountain High School 38.71 41.93 19.35 Bachelor's Degree 42:58 46.129 21.29 Master's Degree 38.7 41.935 19.35 Stop 5: Calculate Chi-square Statistic \$\int \frac{1}{2} \int \text{Cobserved} - \text{Expected} \text{Calculate} \text{Chi-square} \text{Statistic} \text{C}	
	Education level Brach City Mountain High School 38.71 41.93 19.35 Bachelor's Degree 42:58 46.129 21.29 Mayter's Degree 38.7 41.935 19.35 Stop 5:- Calculate Chi-square Statistic X = 5 (Observed - Expected) ² Expected Type ched	
	Education level Brach City Mountain High School 38.71 41.93 19.35 Bachelor's Degree 42:58 46.129 21.29 Mayter's Degree 38.7 41.935 19.35 Stop 5:- Calculate Chi-square Statistic X = 5 (Observed - Expected) ² Expected Type ched	
	Education level Beach City Mountain High School 38.71 41.93 19.35 Bachelor's Degree 42:58 46.129 21.29 Master's Degree 38.7 41.935 19.35 Stop 5: Calculate Chi-square Statistic \$\int \frac{1}{2} \int \text{Cobserved} - \text{Expected} \text{Calculate} \text{Chi-square} \text{Statistic} \text{C}	





magnine you are a Ruality control Manager at a cardy mount activing company. Your company produces bags of axorted cardies and you claim thout the bags contain an equal distribution of four different flavors: "Chocolate" "Strawberry," "Lennon" "Mint" However a consumer group suspects that your bags are not equally distributed and that one of the flavore is indevelopmented. To investigate you randowly solect too cardies from a bag and record the number of each plavor you find you want to test if the observed distribution of flavor; natch your claim.

Observed distribution

Chocolate - 60

Strowberry -50

Lemon -45

Mint - 45

Question

Is there a significant difference blu the observed distribution of randy flavors and the claimed distribution legral distribution of 25% for each flavors) & Perform a Chi-square Goodners of let Test at 5% Significance level.

ANSWER

Step 1:- Set up Hypothesia

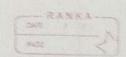
Ho:- The observed distribution matches the daimed distribution of 25% of each flow or)

Aistribution (equal distribution of flavor is as claimed

Ho:- The distribution of flavor is as claimed

H:- Distribution of flavor is not as claimed

	DATE / / STAG
CII	PAGE / /
Stop : Expected Distribution.	
0 11	
Chocolate 5 025 × 200 = 50	
Strawberry = 50	
Cemon > 50	
Mint = 50	
Step 3: Calculate Chi - Square Statistic.	A STATE OF THE STA
S = 3 (Observed - Expected)	•
Expected	
Chocolate 60=100 = 2	4 7
1. 455	Carolin III
Strawberry = 0	5 2 3 1
lemon = 25 = 0.5	- 19) .
lemon = 25 = 0.5	
Min = 0.5	
5°= 3	۸,
Step 4: - Pegree of freedom = 3	.0.
Step 5: - Find withread value from Chi-Square.	table = 7-815
Step 6: - Conclusion	
Crestitical value we exist Null	la bothesis
Step 6: - Conclusion K2 Cuitical value websiteseject Null	Jr



Scoretie: - Suppose you work for a political pellery organizate, and you want to determine whethers there is a significant difference in political party preferences among three difference in political party preferences among three difference age groups: "lours stults" (ages 31-50); and "Senies stults" (ages 31-50); and "Senies stults" (ages 51-50); and "Senies stults" (ages 51-

Datas

Hurds the summarized olata:

- · for "Young Adults" pages 18-30):

 90 individuals prefer farty A

 110 Individuals prefer Party B

 100 Individuals prefer Party C
- · For Middle Aged Adults (ages 31-50)

 . 70 individuals purper Party A

 120 individuals purper Party B

 110 individuals purper Party C
- for "Senior Adults" (ages 51 and above)

 60 individuals prefer Party B

 150 individuals prefer Party C

Ouestion: - 15 there a significant difference in political perty of preferences among the three age groups: "Young that?"

"Middle-Aged Adults", and "Semino - Adults" & Perform a "Chi-Square Test for Homogenity at top significance"

P	TA A		
P		A W SLUER DATE RANKA PAGE PAGE	
	Step 1.	. Ho - The distribution of 2000	
H		Ho - The distribution of political purity preference is same across	all
-		H Distribution of political party is different among	
-		age groups.	1
D	Steh 2		
12	70	Contingency table.	
13		Political Party Profesence	
7		Age Group Pavety A Party B Party C	
10		Young Adults 90 110 100 - 300 Middle-Aged Adults 70 120 110 - 300	
3			
3		Senior Adulty 60 90 150 - 300 220 320 360 900	
3	9ерз.	Chi - Square Statistic	
3		Chi - Square Statistic Experience Statistic	uud
3		000 2 700 2	2
3		Expected value for Young Aduly of Party A= 2449	
			02.
		Age Group Party A Party B Party C	
		Moung Adults 73.33 106.67 120 Middle - Aged Adults 73.33 106.67 120	
3		Made -14ged Adults 73.33 10667 120	
-		Senior Adults 73.33 106.67 120	
		J	
		Expected Value = Row (total) × Column (total)	
		Grand Total	
		$\int_{-\infty}^{\infty} = 3x \left(90 - 73.33 \right)^{2} + \left((10 - 106.67)^{2} + (120 - 100)^{2} \right)$	
		73.33 106.67 120	
2		= 3x (3-789+ 0-1039 + 3.33)	
		2 \$ 21-668	

Steps Conclusion: - Since X 2 > Critical Value we reject null hypothesis Distribution of political party pereference is diff. for age group.